Bootstrap Testing of the Rank of a Matrix via Least Squared Constrained Estimation.

François Portier, Bernard Delyon

Bootstrap Testing of the Rank of a Matrix via Least Squared Constrained Estimation.

François Portier and Bernard Delyon

Abstract In order to test if an unknown matrix has a given rank (null hypothesis), we consider the family of statistics that are minimum squared distances between an estimator and the manifold of fixed-rank matrix. Under the null hypothesis, every statistic of this family converges to a weighted chi-squared distribution. In this paper, we introduce the constrained bootstrap to build bootstrap estimate of the law under the null hypothesis of such statistics. As a result, the constrained bootstrap is employed to estimate the quantile for testing the rank. We provide the consistency of the procedure and the simulations shed light one the accuracy of the constrained bootstrap with respect to the traditional asymptotic comparison. More generally, the results are extended to test if an unknown parameter belongs to a sub-manifold locally smooth. Finally, the constrained bootstrap is easy to compute, it handles a large family of tests and it works under mild assumptions.

Keywords. Rank estimation, Least squared constrained estimation, Bootstrap, Hypothesis testing.

1 Introduction

Let $M_0 \in \mathbb{R}^{p\times H}$ be an unknown matrix (arbitrarily $p \leq H$). To infer about the rank of M_0 with hypothesis testing, the general framework usually considered is the following: there exists an estimator $\widehat{M} \in \mathbb{R}^{p\times H}$ of M_0 such that

$$n^{1/2}(\widehat{M} - M_0) \xrightarrow{\mathrm{d}} W, \qquad \text{with} \qquad \mathrm{vec}\,(W) = \mathcal{N}(0, \Gamma) \tag{1}$$

where $\mathrm{vec}(\cdot)$ vectorizes a matrix by stacking its columns. In the whole paper the hatted quantities are random sequences that depends on the sample number n, all the limit are taken with respect to n. Moreover there exists an estimator $\widehat{\Gamma}$ such that

$$\widehat{\Gamma} \xrightarrow{\mathbb{P}} \Gamma, \tag{2}$$

and in some cases, one may ask that

$$\Gamma \text{ is full rank.} \tag{3}$$

Let d_0 be the rank of M_0 and $m \in \{1, ..., p\}$, we consider the set of hypotheses

$$H_0 \ : \quad d_0 = m \qquad \text{against} \qquad H_1 \ : \quad d_0 > m, \tag{4}$$

Thus d_0 can be estimated the following way: we start by testing $m = 0$, if H_0 is rejected we go a step further $m := m + 1$, if not we stop the procedure and the estimated rank is $\widehat{d} = m$. In this paper, by considering the hypotheses (4) we focus on each step of this procedure.

Many different statistical tests appeared in the literature for this purpose. For instance Cragg and Donald [11] introduced a statistic based on the LU decomposition of $\widehat{M}$, Kleibergen and Paap [19] studied the asymptotic behaviour of some transformation of the singular values of $\widehat{M}$, and Cragg and Donald [12] considered the minimum of a squared distance under rank constraint. In some other fields with similar issues, close ideas have been developed : Bura and Yang [7] examined a Wald type statistic depending on the singular decomposition of $\widehat{M}$ and Cook and Ni [10] also considered the minimum of a squared distance under rank constraint. Although based on different considerations, each of the previous work relies on the test described by (4). For comprehensiveness, in this paper we consider the following three statistics. The first one is introduced by Li [21] as

$$\widehat{\Lambda}_1 = n \sum_{k=m+1}^{p} \widehat{\lambda}_k^2 \tag{5}$$

where $(\widehat{\lambda}_1, ..., \widehat{\lambda}_p)$ are the singular values of $\widehat{M}$ arranged in descending order. Under H_0 and (1), this statistic converges in law to a weighted chi-squared distribution [7]. The main drawback of such a test is that $\widehat{\Lambda}_1$ is not pivotal, i.e. its asymptotic law depends on unknown quantities that are M_0 and Γ. Accordingly the consistency of the associated test requires assumptions (1) and (2). In [7] a standardized version of $\widehat{\Lambda}_1$ is studied with

$$\widehat{\Lambda}_1 = n \operatorname{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)^T [(\widehat{Q}_2 \otimes \widehat{Q}_1) \widehat{\Gamma} (\widehat{Q}_2 \otimes \widehat{Q}_1)]^+ \operatorname{vec}(\widehat{Q}1 \widehat{M} \widehat{Q}2) \tag{6}$$

where M^+ stands for the Moore-Penrose inverse of M and $\widehat{Q}_1$ and $\widehat{Q}_2$ are respectively the orthogonal projectors on the left and right singular spaces associated with the $p - m$ smallest singular values of $\widehat{M}$. The authors proved that under H_0, if (1) and (2) hold, the Wald-type statistic $\widehat{\Lambda}_2$ is asymptotically chi-squared distributed. Besides, [12] and [10] proposed a constrained estimator by minimizing a squared distance under a fixed-rank constraint as

$$\widehat{\Lambda}_3 = n \min_{\operatorname{rank}(M)=m} \operatorname{vec}(\widehat{M} - M)^T \widehat{\Gamma}^{-1} \operatorname{vec}(\widehat{M} - M), \tag{7}$$

which is also asymptotically chi-squared distributed under H_0, assuming (1), (2) and (3). We will refer the minimum discrepancy approach. Although the statistics $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ have the convenience of being pivotal, they both require the inversion of a large matrix and this may cause robustness problems when the sample number is not large enough. For $\alpha \in]0, 1[$ and under the relevant assumptions, each of these statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$, is consistent at level α in testing (4), i.e. the level goes to $1 - \alpha$ and the power goes to 1 as n goes to ∞.

Nevertheless the estimation of the quantile is difficult because either the asymptotic distribution depends on the data (non pivotality represented by $\widehat{\Lambda}_1$), or the true distribution may be quite different than the asymptotic one (slow rates of convergence represented by $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$). The objective of the paper is to propose a bootstrap method for quantile estimation in this context.

An important remark which instigates the sketch of the paper is that all the previous statistics share the form

$$\widehat{\Lambda} = n \|\widehat{B} \operatorname{vec}(\widehat{M} - \widehat{M}_c)\|^2 \quad \text{with} \quad \widehat{M}_c = \underset{\operatorname{rank}(M)=m}{\operatorname{argmin}} \|\widehat{A} \operatorname{vec}(\widehat{M} - M)\|^2 \tag{8}$$

where $\| \cdot \|$ is the Euclidean norm, $\widehat{A} \in \mathbb{R}^{pH \times pH}$, $\widehat{B} \in \mathbb{R}^{pH \times pH}$. The values of $\widehat{A}$ and $\widehat{B}$ corresponding to the statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ are summarized in the Table 1 (See Section 2 for the details).

We refer to traditional testing (resp. bootstrap testing) when the statistic is compared to its asymptotic quantile (resp. bootstrap quantile). The bootstrap test is said to be consistent

	$\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$
$\widehat{A}$	I	I	$\widehat{\Gamma}^{-1/2}$
$\widehat{B}$	I	$[(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)]^{+1/2}$	$\widehat{\Gamma}^{-1/2}$

Table 1: Values of $\widehat{A}$ and $\widehat{B}$ in (8) for $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$.

at level α if

$$\mathbb{P}_{H_0}\left(\widehat{\Lambda} > \widehat{q}(\alpha)\right) \longrightarrow 1-\alpha \qquad \text{and} \qquad \mathbb{P}_{H_1}\left(\widehat{\Lambda} > \widehat{q}(\alpha)\right) \longrightarrow 1, \tag{9}$$

where $\widehat{q}(\alpha)$ is the quantile of level α calculated by bootstrap. The advantage of bootstrap testing is its high level of accuracy under H_0 with respect to traditional testing. This fact is emphasized by considering the two possibilities: when the statistic is pivotal and when the asymptotic law of the statistic depends on unknown quantities. First, as highlighted by Hall [15], when the statistic is pivotal, under some conditions the gap between the distribution of the statistic and its bootstrap distribution is $O_{\mathbb{P}}(n^{-1})$. Since the normal approximation leads to a difference $O(n^{1/2})$, the bootstrap enjoys a better level of accuracy. Secondly if the asymptotic law of the statistic is unknown, the bootstrap appears even more as a convenient alternative because it avoids its estimation. In [17], Hall and Wilson give two advices for the use of the bootstrap testing:

A) Whatever the sample is under H_0 or H_1, the bootstrap estimates the law of the statistic under H_0.

B) The statistic is pivotal.

The first guideline is the most crucial because if it fails it may lead to inconsistency of the test. The second guideline aims at improving the accuracy of the test by taking full advantage of the accuracy of the bootstrap. In this paper we propose a new procedure for bootstrap testing in least square constraint estimation (LSCE) (estimators as (8) are particular cases), called constrained bootstrap (CS bootstrap). More precisely, the CS bootstrap aims at testing whether a parameter belongs or not to a submanifold and so generalised the test (4). Our main result is the consistency of the CS bootstrap under mild conditions. As a consequence we provide a consistent bootstrap testing procedure for testing (4) with the statistic $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. For the sake of clarity, we address the CS bootstrap in the next section. Section 3 is dedicated to rank estimation with special interest to the bootstrap of the statistic $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. Finally, the last section emphasizes the accuracy of the bootstrap in rank estimation by providing a simulation study in sufficient dimension reduction (SDR). Accordingly, the sketch of the paper is as follows:

- The CS bootstrap in LSCE
- Bootstrap testing procedure for $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$
- Application to SDR

2 The constrained bootstrap for LSCE and hypothesis testing

Because of (8) LSCE has a central place in the paper. Moreover since LSCE intervenes in many statistical fields as M-estimation or hypothesis testing, this section is independent from the rest of the paper.

2.1 LSCE

Let $\theta_0 \in \mathbb{R}^p$ be called the parameter of interest, and let $\widehat{\theta} \in \mathbb{R}^p$ be an estimator of θ_0. We define the constrained estimator of θ_0 as

$$\widehat{\theta}_c = \operatorname*{argmin}_{\theta \in \mathcal{M}} (\widehat{\theta} - \theta)^T \widehat{A} (\widehat{\theta} - \theta), \tag{10}$$

where $\mathcal{M}$ is a submanifold of $\mathbb{R}^p$ with co-dimension q, and $\widehat{A} \in \mathbb{R}^{p \times p}$. The constrained statistic is defined as

$$\widehat{\Lambda} = n(\widehat{\theta} - \widehat{\theta}_c)^T \widehat{B} (\widehat{\theta} - \widehat{\theta}_c). \tag{11}$$

where $\widehat{B} \in \mathbb{R}^{p \times p}$. Note that if $\widehat{A}$ is full rank, the unique minimizer of (10) without constraint is $\widehat{\theta}$, hence it could be understood as the unconstrained estimator. We introduce now the notion of nonsingular point in $\mathcal{M}$. This one is needed to express the Lagrangian first order condition of the optimization (10). For any function $g = (g_1, \dots, g_p) : \mathbb{R}^p \to \mathbb{R}^q$, define its Jacobian as $J_g = (\nabla g_1, ..., \nabla g_q)$, where ∇ stands for the gradient operator.

Definition 1. *We say that θ is $\mathcal{M}$-nonsingular if $\theta \in \mathcal{M}$ and if there exists a neighbourhood V and a function $g : \mathbb{R}^p \to \mathbb{R}^q$ continuously differentiable on V with $J_g(\theta)$ full rank such that*

$$V \cap \mathcal{M} = \{g = 0\}.$$

As a consequence any point of a submanifold locally smooth is nonsingular, e.g. any matrix with rank m is a nonsingular point in the submanifold $\operatorname{rank}(M) = m$. We prove in Proposition 2 that if θ_0 is $\mathcal{M}$-nonsingular, $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \Delta)$ and $\widehat{B} = \widehat{A} \xrightarrow{\mathbb{P}} A$ is full rank, then we have

$$\widehat{\Lambda} \xrightarrow{\mathrm{d}} \sum_{k=1}^{p} \nu_k W_k^2, \tag{12}$$

where the W_k's are i.i.d. Gaussian random variables and the ν_k's are the eigenvalues of the matrix $\Delta^{1/2} J_g(\theta_0)^T (J_g(\theta_0) A^{-1} J_g(\theta_0)^T)^{-1} J_g(\theta_0) \Delta^{1/2}$. Especially, the case $A = \Delta^{-1}$ is interesting because $\widehat{\Lambda}$ is asymptotically chi-squared distributed with q degrees of freedom. Otherwise, if $\theta_0 \notin \mathcal{M}$, $\widehat{\Lambda}$ goes to infinity in probability. Those facts shed light on a consistent testing procedure based on LSCE with the hypotheses

$$H_0 \ : \quad \theta_0 \in \mathcal{M} \qquad \text{against} \qquad H_1 \ : \quad \theta_0 \notin \mathcal{M} \tag{13}$$

and the decision rule to reject H_0 if $\widehat{\Lambda}$ is larger than a quantile of its asymptotic law. Accordingly the previous framework can be seen as an extension of the Wald test statistic which handles the simple hypothesis $\theta_0 = \theta$ with the statistic $(\widehat{\theta} - \theta)^T \Delta^{-1} (\widehat{\theta} - \theta)$.

2.2 The bootstrap in LSCE

Since LSCE is a particular case of estimating equation, we review the bootstrap literature with two principal directions: estimating equation and hypothesis testing. For clarity we alleviate the framework in this section: let $X_1, \cdots, X_n$ be an i.i.d. sequence of real random variables with law P, define $\gamma = \operatorname{var}(X_1)$, $\widehat{\gamma} = \overline{(X - \overline{X})^2}$, we put $\theta_0 = \mathbb{E}[X_1]$, $\widehat{\theta} = \overline{X}$, and $A = B = \gamma^{-1}$ where $\overline{\cdot}$ stands for the empirical mean.

The original bootstrap was introduced in [14] in the following way. Let $X_1^*, \dots, X_n^*$ be an i.i.d. sequence of real random variables with law $\widehat{P} = n^{-1} \sum_{i=1}^n \delta_{X_i}$, define $\theta^* = \overline{X^*}$, the distribution of $\sqrt{n}(\theta^* - \widehat{\theta})$ conditionally on the sample, that we call the bootstrap distribution, is "close" to the distribution of $\sqrt{n}(\widehat{\theta} - \theta_0)$, that we call the true distribution (in the rest of the paper we just say "conditionally" instead of "conditionally on the sample"). For instance, it is shown in [23] that the bootstrap distribution converges weakly to the true distribution almost surely. One says that $\sqrt{n}(\theta^* - \widehat{\theta})$ bootstraps $\sqrt{n}(\widehat{\theta} - \theta_0)$ and we will write

$$\mathcal{L}_\infty(n^{1/2}(\theta^* - \widehat{\theta})|\widehat{P}) = \mathcal{L}_\infty(n^{1/2}(\widehat{\theta} - \theta_0)) \quad \text{a.s.},$$

where $\mathcal{L}_\infty(\cdot)$ and $\mathcal{L}_\infty(\cdot|\widehat{P})$ both mean the asymptotic laws with the difference that the later is conditional on the sample. Equivalently, one has for every $x \in \mathbb{R}$, $\mathbb{P}(\sqrt{n}(\theta^* - \widehat{\theta}) \leq x|\widehat{P}) \overset{\text{a.s.}}{\to} \mathbb{P}(\sqrt{n}(\widehat{\theta} - \theta_0) \leq x)$, but the use of the bootstrap is legitimate by a more general results stated in [15], which says that

$$|\mathbb{P}(n^{1/2}(\theta^* - \widehat{\theta})/\gamma^* \leq x|\widehat{P}) - \mathbb{P}(n^{1/2}(\widehat{\theta} - \theta_0)/\widehat{\gamma} \leq x)| = O_{\mathbb{P}}(n^{-1}) \tag{14}$$

with $\gamma^* = \overline{(X^* - \overline{X^*})^2}$, provided that P is non-lattice. Besides, one has

$$|\mathbb{P}(n^{1/2}(\widehat{\theta} - \theta_0)/\widehat{\gamma} \leq x) - \Phi(x)| = O_{\mathbb{P}}(n^{-1/2}),$$

where Φ is the cumulative distribution function (c.d.f.) of the standard normal law. Variations of Efron's resampling plan are proposed in [2] under the name of weighted bootstrap. For a complete introduction about the bootstrap we refer to [15]. We now present three different bootstrap techniques related to LSCE[1].

(i) The classical bootstrap (C bootstrap)

The literature about the bootstrap in Z and M-estimation, see respectively [9] and [1], is based on the following principle: if $\theta_M = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \mathbb{E}[\phi(X,\theta)]$ is estimated by $\widehat{\theta}_M = \underset{\theta\in\Theta}{\operatorname{argmin}}\ \frac{1}{n}\sum_{i=1}^n \phi(X_i,\theta)$ where Θ is an open set, then the bootstrap of $\sqrt{n}(\widehat{\theta}_M - \theta_M)$ is carried out by the quantity $\sqrt{n}(\theta^*_M - \widehat{\theta}_M)$ with

$$\theta^*_M = \underset{\theta\in\Theta}{\operatorname{argmin}}\ n^{-1}\sum_{i=1}^n w_i\phi(X_i,\theta), \tag{15}$$

where (w_i) is a sequence of random variables. The particular case where the vector $(w_1,\dots,w_n)$ is distributed as $\text{mult}(n,(n^{-1},\dots,n^{-1}))$ leads to a direct application of original Efron's bootstrap to M-estimation. Since such a bootstrap has been extensively studied, we refer to the C bootstrap. To the knowledge of the authors, the C bootstrap when Θ has empty interior has not been studied yet. Nevertheless one may sight its bad behaviour for the test of equal mean $H_0: \ \theta_0 = \mu$. The associated least squared constrained statistic

$$n\widehat{\gamma}^{-1}(\widehat{\theta} - \mu)^2,$$

is indeed the score statistic associated to the M-estimator with $\phi(x,\theta) = \widehat{\gamma}^{-1}(x-\theta)^2$ and $\Theta = \{\mu\}$. Clearly the C bootstrap through $n\gamma^{*-1}(\theta^* - \mu)^2$ does not work because of its bad behaviour under H_1 for instance. In this case it is better to use

$$n\gamma^{*-1}(\theta^* - \widehat{\theta})^2,$$

[1] A bootstrap with a Delta-method approach (see [23], chapter 23, Theorem 5) fails because $x \to \min_{\|\theta\|=1}\|x-\theta\|$ is not continuously differentiable on the unit circle.

but it cannot handle the cases of more involved hypotheses[2]. Whereas the C bootstrap is not really connected with hypothesis testing, the two following bootstrap procedures are more related to the present work.

(ii) The biaised bootstrap (B bootstrap)

The B bootstrap is introduced in [16] and is directly motivated by hypothesis testing. The original idea of their work is to re-sample with respect to the distribution $\widehat{P}_b = \sum_{i=1}^n \omega_i \delta_{X_i}$, where the ω_i's maximize

$$\sum_{i=1}^n \log(\omega_i) \quad \text{under the constraints} \quad \begin{array}{c} \frac{1}{n}\sum_{i=1}^n \omega_i X_i = \mu \\ \sum_{i=1}^n \omega_i = 1 \end{array} . \tag{16}$$

Since the ω_i's minimize the Kulback-Leibler distance between $\widehat{P}$ and $\widehat{P}_b$, one can see the resulting distribution as the closest to the original one satisfying the mean constraint. The authors presented interesting results for the test of equal mean $\theta_0 = \mu$, essentially the bootstrap statistic $n\gamma^{*-1}(\theta_b^* - \mu)^2$, with $\theta_b^* = \overline{X_b^*}$, $X_{b,i}^*$ sampled from $\widehat{P}_b$, has a chi-squared limiting distribution either H_0 or H_1 is assumed. As a result both guidelines (A) and (B) are checked. They go further by showing that the B bootstrap outclasses the asymptotic normal approximation for quantile estimation in the sense that $|\widehat{q}(\alpha) - q_n(\alpha)| = O_{\mathbb{P}}(n^{-1})$ whereas $|q_n(\alpha) - q_\infty(\alpha)| = O(n^{-1/2})$, where q_∞, q_n and $\widehat{q}_n$ are the quantile functions of the standard normal distribution, the statistic $n\widehat{\gamma}^{-1}(\widehat{\theta} - \mu)^2$ under H_0 and the bootstrapped statistic, respectively. Although the B bootstrap matches the context of hypothesis testing, it has been designed to handle the particular test of equal mean. To the knowledge of the authors the study of the B bootstrap has not been extended to other tests. Facing (16), the main drawback of the B bootstrap deals with algorithmic difficulties. Indeed when the constraint becomes more involved, solving (16) is more difficult. As a result it is not sure that this method could handle other situations such as fixed-rank constraints.

(iii) The estimating function bootstrap (EF bootstrap)

Now $X_i \in \mathbb{R}^p$. Some other ideas about the bootstrap of the Z-estimators can be found in [20] and [18], and can be summarized as follows. Considering the score statistic $\widehat{S} = \sqrt{n}\sum_{i=1}^n \frac{\partial \phi}{\partial \theta}(X_i, \theta_0)$, [18] showed that it could be bootstrapped by

$$S^* = n^{-1/2} \sum_{i=1}^n w_i \frac{\partial \phi}{\partial \theta}(X_i, \widehat{\theta}),$$

where (w_i) is a sequence of random variables. This bootstrap is called the EF bootstrap and revealed nice computational properties. Moreover the authors argued for its use in quantile estimation in order to test if $g(\theta_0) = 0$, where $g : \mathbb{R}^p \to \mathbb{R}^q$ is the constraint function, by recommending essentially to use $S^{*T} J_g(\widehat{\theta})^T \left(J_g(\widehat{\theta})\gamma^* J_g(\widehat{\theta})^T\right)^{-1} J_g(\widehat{\theta}) S^*$. Applying it to the least squared context $\phi(x, \theta) = \|\widehat{\gamma}^{-1/2}(x - \theta)\|^2$, the EF bootstrap is carried out by

$$n(\theta^* - \widehat{\theta})^T J_g(\widehat{\theta})^T \left(J_g(\widehat{\theta})\gamma^* J_g(\widehat{\theta})^T\right)^{-1} J_g(\widehat{\theta})(\theta^* - \widehat{\theta}).$$

Although it verifies both guidelines (A) and (B) (see the article for details), one can see that the good behaviour of such an approach is more based on the rank deficiency

[2]We refer to [17] for a study of this bootstrap in order to test $\theta_0 = \mu$.

of $J_g(\widehat{\theta})$ than on the bootstrap of $\sqrt{n}(\widehat{\theta}-\widehat{\theta}_c)$. Indeed $\sqrt{n}(\theta^*-\widehat{\theta})$ bootstraps the non constrained estimator $\sqrt{n}(\widehat{\theta}-\theta_0)$. Then as the authors noticed, it is first of all a bootstrap of the Wald-type statistic $n\widehat{S}^T J_g(\theta_0)^T \left(J_g(\theta_0)\widehat{\gamma}J_g(\theta_0)^T\right)^{-1} J_g(\theta_0)\widehat{S}$ which has fortunately the same asymptotic law than the targeted one. This may induce some loss in accuracy. Moreover, it requires the knowledge of the function J_g which is not the case for fixed rank constraints where the g depends on the limit M_0 (see Remark 1 for some details).

Essentially both (i) and (ii) provide a bootstrap for testing simple hypotheses. The EF bootstrap proposed in (iii) extends this limited scope by including tests of the form $g(\theta_0)=0$ where g is known. Nevertheless it does not handle the test (4) as it is highlighted by the following remark.

Remark 1. Testing (4) with $\widehat{\Lambda}_3$ results in an optimization with the constraint $\operatorname{rank}(M)=m$. Since the subspace of fixed rank matrices is a submanifold locally smooth with co-dimension $(p-d)(H-d)$, at every point M, there exists a neighbourhood V and a $\mathcal{C}^\infty$ function $g: V \to \mathbb{R}^{(p-d)(H-d)}$ such that $V \cap \{\operatorname{rank}(M)=m\}=\{g=0\}$ and $J_g(M)$ has full rank. Moreover, we have

$$\|\Gamma^{-1/2}\operatorname{vec}(\widehat{M}_c-M_0)\| \leq 2\|\Gamma^{-1/2}\operatorname{vec}(\widehat{M}-M_0)\|.$$

If now (1) holds, the right-hand side term goes to 0 in probability and $\widehat{M}_c \xrightarrow{\mathbb{P}} M_0$. As a consequence, if Γ is invertible, for any neighbourhood of M_0, from a certain rank, $\widehat{M}_c$ belongs to it with probability 1. Then under H_0 since M_0 has rank m the constrained estimator has the expression

$$\widehat{M}_c = \underset{g(M)=0}{\operatorname{argmin}}\|\Gamma^{-1/2}\operatorname{vec}(\widehat{M}_c-M)\|,$$

with g depending on M_0. Unfortunately we do not know neither g nor $J_g(M_0)$. This entails some problems relating to the later approach.

2.3 The constrained bootstrap

The CS bootstrap is introduced in order to solve all the issues we have raised through the previous little review which are essentially: computational difficulties and small scope of the existing methods. The CS bootstrap targets an estimation $\widehat{q}(\alpha)$ of the quantile under H_0 of $\widehat{\Lambda}$. The consistency of the procedure, i.e. (9), forms the main result about the CS bootstrap. Another important issue which occurs beforehand in the section is the bootstrap of the law of

$$n^{1/2}(\widehat{\theta}_c-\theta_0) \quad \text{under } H_0.$$

Basically, we show that a bootstrap of the unconstrained estimator $\sqrt{n}(\widehat{\theta}-\theta_0)$ allows a bootstrap of the constrained estimator $\sqrt{n}(\widehat{\theta}_c-\theta_0)$ under H_0. We point out that the CS bootstrap heuristic is rather different than the C and EF bootstrap. Otherwise it shares the idea to "reproduce" H_0 even if H_1 is realized with the B bootstrap. Assuming that we can bootstrap $\sqrt{n}(\widehat{\theta}-\theta_0)$, the CS bootstrap calculation of the statistic is realized as follows:

The CS bootstrap procedure
Compute

$$\theta_0^* = \widehat{\theta}_c + n^{-1/2} W^*, \qquad \text{with} \quad \mathcal{L}_\infty(W^*|\widehat{P}) = \mathcal{L}_\infty(n^{1/2}(\widehat{\theta} - \theta_0)) \quad \text{a.s.}, \tag{17}$$

where the simulation of W^* can be done by a standard bootstrap procedure[3]. Calculate

$$\theta_c^* = \operatorname*{argmin}_{\theta \in \mathcal{M}} \, (\theta_0^* - \theta)^T A^* (\theta_0^* - \theta), \qquad \text{and} \qquad \Lambda^* = n(\theta_0^* - \theta_c^*)^T B^* (\theta_0^* - \theta_c^*), \tag{18}$$

where $A^* \in \mathbb{R}^{p \times p}$ and $B^* \in \mathbb{R}^{p \times p}$ [4].

Intuitively, this choice appears natural because θ_0^* equals $\widehat{\theta}_c$ plus a small perturbation going to 0. Accordingly θ_0^* is somewhat reproducing the behaviour of $\widehat{\theta}$ under H_0, especially because W^* has the right asymptotic variance. As we should notice, A^* and B^* could be chosen as $\widehat{A}$ and $\widehat{B}$ but this is not the best choice in practice. As it is highlighted in (14), we should normalized by the associated bootstrap quantities (e.g. the variance computed on the bootstrap sample). The following lemma gives a first order decomposition of the bootstrap law $\sqrt{n}(\theta_c^* - \widehat{\theta}_c)$ under mild conditions. The following lemma is proved in the Appendix.

Lemma 1. *Let $\mathcal{M}$ be a submanifold. Assume there exists $\widehat{\theta}_c \in \mathcal{M}$ and θ_c a $\mathcal{M}$-nonsingular point such that $\widehat{\theta}_c \xrightarrow{\text{a.s.}} \theta_c$. If moreover $\mathcal{L}_\infty(\sqrt{n}(\theta_0^* - \widehat{\theta}_c)|\widehat{P})$ exists a.s. and conditionally a.s. $A^* \xrightarrow{\mathbb{P}} A$ is full rank, then we have conditionally a.s.*

$$n^{1/2}(\theta_c^* - \widehat{\theta}_c) = (I - P) n^{1/2}(\theta_0^* - \widehat{\theta}_c) + o_{\mathbb{P}}(1),$$

with $P = A^{-1} J_g^T(\theta_c)(J_g(\theta_c) A^{-1} J_g^T(\theta_c))^{-1} J_g(\theta_c)$.

Note that if θ_0 is $\mathcal{M}$-nonsingular and $\mathcal{L}_\infty(\sqrt{n}(\widehat{\theta} - \theta_0)|\widehat{P})$ exists, we can apply Lemma 1 with $\widehat{\theta}_c = \theta_c = \theta_0$. This gives the following proposition:

Proposition 2. *Let $\mathcal{M}$ be a submanifold. Assume that $\mathcal{L}_\infty(\sqrt{n}(\widehat{\theta} - \theta_0)|\widehat{P})$ exists with θ_0 $\mathcal{M}$-nonsingular. Assume also that $\widehat{A} \xrightarrow{\mathbb{P}} A$ is full rank, then we have*

$$n^{1/2}(\widehat{\theta}_c - \theta_0) = (I - P) n^{1/2}(\widehat{\theta} - \theta_0) + o_{\mathbb{P}}(1),$$

with $P = A^{-1} J_g^T(\theta_0)(J_g(\theta_0) A^{-1} J_g^T(\theta_0))^{-1} J_g(\theta_0)$.

Proposition 2 leads easily to (12) and extends classical results [6] about constrained estimators with constraint $\{g = 0\}$ to manifold type constraints. Besides statements of Lemma 1 and Proposition 2 together explain the preceding definition of θ_0^* in (17). They also lead to the following theorem.

Theorem 3. *Let $\mathcal{M}$ be a submanifold. Assume that $\widehat{\theta} \xrightarrow{\text{a.s.}} \theta_0$ with θ_0 $\mathcal{M}$-nonsingular and $\widehat{A} \xrightarrow{\mathbb{P}} A$ hold. If moreover (17) holds and conditionally a.s. $A^* \xrightarrow{\mathbb{P}} A$ is full rank, then we have*

$$\mathcal{L}_\infty(n^{1/2}(\theta_c^* - \widehat{\theta}_c)|\widehat{P}) = \mathcal{L}_\infty(n^{1/2}(\widehat{\theta}_c - \theta_0)) \quad \text{a.s.}.$$

[3] The bootstrap procedure to get W^* is not specified because it depends on $\widehat{\theta}$. For instance, if $\widehat{\theta}$ is a mean over some i.i.d. random variables, one can use the Efron's traditional bootstrap and if $\widehat{\theta}$ is a M-estimator, one should use a bootstrap as detailed by equation (15).

[4] Assumptions about A^* and B^* are provided further in the statements of the propositions.

Essentially, Theorem 3 is an application of Lemma 1 under H_0, indeed as we seen in the proof of Lemma 1, equation (24), the assumption $\widehat{\theta} \stackrel{\text{a.s.}}{\to} \theta_0 \in \mathcal{M}$ implies that $\widehat{\theta}_c \stackrel{\text{a.s.}}{\to} \theta_c$. Nevertheless under H_1 nothing guarantee such a convergence (see Example 1 below). Roughly speaking, asking for an equality in law under H_1 as in Theorem 3 may be too much to ask. However as stated in the following theorem we do not require that $\widehat{\theta}_c$ converges a.s. to a constant to provide that the power of the corresponding test goes to 1. This leads to the consistency of the CS bootstrap for hypothesis testing. For the statement of the consistency theorem, we need to define the quantile function of the bootstrap statistic

$$\widehat{q}(\alpha) = \inf \{x \ : \ \widehat{F}(x) \geq 1 - \alpha\},$$

where $\widehat{F}$ is the c.d.f. of Λ^* conditionally on the sample.

Theorem 4. *Let $\mathcal{M}$ be a manifold. Assume that $\widehat{\theta} \stackrel{\text{a.s.}}{\to} \theta_0$ with θ_0 $\mathcal{M}$-nonsingular under H_0. We assume also that $\widehat{A} \stackrel{\mathbb{P}}{\to} A$ is full rank, $\widehat{B} \stackrel{\mathbb{P}}{\to} B$. If moreover $\mathcal{L}_\infty(\sqrt{n}(\theta_0^* - \widehat{\theta}_c)|\widehat{P}) = \mathcal{L}_\infty(\sqrt{n}(\widehat{\theta} - \theta_0))$ a.s. has a density, and conditionally a.s. $A^* \stackrel{\mathbb{P}}{\to} A$, $B^* \stackrel{\mathbb{P}}{\to} B$, then we have*

$$\mathbb{P}_{H_0}(\widehat{\Lambda} > \widehat{q}(\alpha)) \longrightarrow 1 - \alpha, \quad \text{and} \quad \mathbb{P}_{H_1}(\widehat{\Lambda} > \widehat{q}(\alpha)) \longrightarrow 1.$$

In other words, the test described in (13) with statistic $\widehat{\Lambda}$ and CS bootstrap calculation of quantile is consistent.

We provide the following example under H_1, where $\widehat{\theta}_c$ does not converge to a constant in probability. Although we cannot get the conclusion of Theorem 3, the least squared constrained statistic still converges in distribution.

Example 1. Let $(X_i)_{i \in \mathbb{N}}$ be a i.i.d. sequence such that $X_1 \stackrel{\text{d}}{=} \mathcal{N}(0,1)$. Define $\widehat{\theta} = \overline{X}$, and $H_0 : \ \theta_0^2 = 1$. Clearly H_0 does not hold and naturally the statistic $n \min\limits_{\theta^2=1} \|\widehat{\theta} - \theta\|^2$ goes to infinity in probability. One can find that $\widehat{\theta}_c = \operatorname{sign}(\overline{X})$ which does not converge. Since

$$\theta_c^* = \operatorname*{argmin}_{\theta^2=1} \|\theta_0^* - \theta\|^2 \quad \text{and} \quad \theta_0^* = \widehat{\theta}_c + n^{-1/2} W^*,$$

we get that $\theta_c^* = \widehat{\theta}_c$ a.s. and naturally, we do not have the asymptotic given by Theorem 3. Besides, the convergence to a chi-squared distribution holds for the quantity $n \min\limits_{\theta^2=1} \|\theta_0^* - \theta\|^2$.

3 Rank estimation with hypothesis testing

In this section through a review of the literature about rank estimation, we apply the results obtained in section 2.1 to provide a consistent bootstrap procedure for the test described by (4) associated with the statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. We define $q_0 = p - d_0$ the dimension of the kernel of M_0^T. We denote by $(\lambda_1, ..., \lambda_p)$ the singular values of M_0 arranged in descending order and we write the SVD of M_0 as

$$M_0 = (U_1 U_0) \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_0^T \end{pmatrix},$$

with $U_1 \in \mathbb{R}^{p \times d_0}$, $U_0 \in \mathbb{R}^{p \times q_0}$, $V_1 \in \mathbb{R}^{H \times d_0}$, $V_0 \in \mathbb{R}^{H \times q_0}$, and $D_1 = \operatorname{diag}(\lambda_1, ..., \lambda_{d_0})$. For $m \in \{1, \cdots, p\}$, we note $q = p - m$ and we write the SVD of $\widehat{M}$ as

$$\widehat{M} = (\widehat{U}_1 \widehat{U}_0) \begin{pmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_0 \end{pmatrix} \begin{pmatrix} \widehat{V}_1^T \\ \widehat{V}_0^T \end{pmatrix},$$

with $\widehat{U}_1 \in \mathbb{R}^{p\times m}$, $\widehat{U}_0 \in \mathbb{R}^{p\times q}$, $\widehat{V}_1 \in \mathbb{R}^{H\times m}$, $\widehat{V}_0 \in \mathbb{R}^{H\times q}$, $\widehat{D}_1 = \mathrm{diag}(\widehat{\lambda}_1, ..., \widehat{\lambda}_m)$ and $\widehat{D}_0 = \mathrm{diag}(\widehat{\lambda}_{m+1}, ..., \widehat{\lambda}_p)$. We also introduce the orthogonal projectors

$$Q_1 = I - P_1 = U_0 U_0^T,\ Q_2 = I - P_2 = V_0 V_0^T,\ \widehat{Q}_1 = I - \widehat{P}_1 = \widehat{U}_0 \widehat{U}_0^T \text{ and } \widehat{Q}_2 = I - \widehat{P}_2 = \widehat{V}_0 \widehat{V}_0^T.$$

Whereas the link between $\widehat{\Lambda}_3$ and LSCE is evident, the one conecting $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$ to LSCE relies on the following classical lemma, whose proof is avoided.

Lemma 5. *Let $\widehat{M} \in \mathbb{R}^{p\times H}$, it holds that*

$$\underset{\mathrm{rank}(M)=m}{\mathrm{argmin}} \ \|\widehat{M} - M\|_F^2 = \widehat{P}_1 \widehat{M} \widehat{P}_2, \quad \textit{and} \quad \|\widehat{M} - \widehat{P}_1 \widehat{M} \widehat{P}_2\|_F^2 = \sum_{k=m+1}^{p} \widehat{\lambda}_k^2,$$

where $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_p$ are the singular values of $\widehat{M}$ arranged in descending order, and $\widehat{P}_1$ and $\widehat{P}_2$ are orthogonal right and left singular projectors of $\widehat{M}$ associated with $\widehat{\lambda}_1, \ldots, \widehat{\lambda}_m$.

Note that in the previous lemma, $\widehat{P}_1$ and $\widehat{P}_2$ are uniquely determined if and only if $\widehat{\lambda}_m \neq \widehat{\lambda}_{m+1}$.

3.1 Nonpivotal statistic

As stated in the introduction, the statistic $\widehat{\Lambda}_1 = n \sum_{k=m+1}^{p} \widehat{\lambda}_k^2$ can be used to arbitrate between the hypotheses of (4). Basically, if $H_0 :\ d_0 = m$ is realized, all the eigenvalues of the sum goes to 0 and $\widehat{\Lambda}_1$ has a weighted chi-squared limiting distribution. Otherwise, at least one eigenvalue converges in probability to a positive number and for any $A > 0$, $\mathbb{P}(\widehat{\Lambda}_1 > A) \longrightarrow 1$. The following proposition describes the asymptotic behaviour of $\widehat{\Lambda}_1$[5]. It was stated in [8] and some recent extension can be found in [7]. Our statement goes further because we are also concerned about the estimation of the asymptotic law of $\widehat{\Lambda}_1$, i.e. the estimation of the weights that intervenes in the weighted chi-squared asymptotic law. Besides, the proof we give in the Appendix is quite simple[6].

Proposition 6. *Under H_0, if (1) holds we have*

$$\widehat{\Lambda}_1 \xrightarrow{d} \sum \nu_k W_k^2$$

where the ν_k's are the eigenvalues of the matrix $(Q_2 \otimes Q_1)\Gamma(Q_2 \otimes Q_1)$ and the W_k's are i.i.d. standard Gaussian variables. If moreover (2) holds, we have

$$(\widehat{\nu}_1, ..., \widehat{\nu}_{pH}) \xrightarrow{\mathbb{P}} (\nu_1, ..., \nu_{pH}),$$

where the $\widehat{\nu}_k$'s are the eigenvalues of the matrix $(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)$.

Remark 2. Unlike Theorem 1 in [8] or Theorem 1 in [7], we prefer to state this theorem with the quantities Q_1 and Q_2 rather than with U_0 and V_0. Because we do not assume that the kernel of M has dimension 1, the vectors that form U_0 or V_0 are not unique because vector spaces with dimension larger than 2 have an infinite number of basis. As a consequence it does not make sense to estimate either U_0 or V_0. To characterize convergence of spaces, a suitable object is their associated orthogonal projectors.

[5] A similar proposition can be stated applying Proposition 12. Following this way, the asymptotic depends on g which is difficult to estimate for rank constraints (see Remark 1).

[6] We no longer need the results of [13] about the asymptotic behaviour of singular values.

In general, we do not know the asymptotic distribution of $\widehat{\Lambda}_1$ because it depends on $(Q_2 \otimes Q_1)\Gamma(Q_2 \otimes Q_1)$. On the first hand, one can estimate consistently this matrix to get an approximation of the law of $\widehat{\Lambda}_1$ under H_0. Some conditions providing the consistency of the estimation are stated in Proposition 6. On the other hand, one can apply the CS bootstrap to estimate the quantile of $\widehat{\Lambda}_1$ in order to test. The main advantage of such an approach is that we no longer need to have a consistent estimator of Γ so that (2) is not needed anymore. Following section 2.1 and by using Lemma 5, we define

$$M_0^* = \widehat{P}_1 \widehat{M} \widehat{P}_2 + n^{-1/2} W^* \qquad \text{with} \quad W^* | \widehat{P} \xrightarrow{\mathrm{d}} W \quad \text{a.s.}, \tag{19}$$

with W defined in (1). Accordingly, we introduce the CS bootstrap statistic

$$\Lambda_1^* = n \sum_{k=m+1}^{p} \lambda_k^{*2},$$

with $\lambda_{m+1}^*, ..., \lambda_p^*$ the smallest singular values of M^*. The following proposition is a straightforward application of Theorem 4 with the submanifold $\{\mathrm{rank}(M) = m\}$.

Proposition 7. *If (1), (19) and $\widehat{M} \xrightarrow{\text{a.s.}} M_0$ hold, then the test described in (4) with the statistic $\widehat{\Lambda}_1$ and calculation of quantile with $\widehat{\Lambda}_1^*$ is consistent.*

3.2 Wald-type statistic

The Wald-type statistic $\widehat{\Lambda}_2 = \mathrm{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)^T [(\widehat{Q}_2 \otimes \widehat{Q}_1) \widehat{\Gamma} (\widehat{Q}_2 \otimes \widehat{Q}_1)]^+ \mathrm{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)$ has been introduced in [7] to get a pivotal statistic[7]. They obtained the following theorem for which we provide a different proof in the appendix.

Proposition 8. *If (1) and (2) hold, we have*

$$\widehat{\Lambda}_2 \xrightarrow{\mathrm{d}} \chi_s^2,$$

with $s = \min(\mathrm{rank}(\Gamma), (p-d)(H-d))$.

Following (18), we define the associated bootstrap statistic by

$$\widehat{\Lambda}_2^* = \mathrm{vec}(Q_1^* M_0^* Q_2^*)^T [(Q_2^* \otimes Q_1^*) \Gamma^* (Q_2^* \otimes Q_1^*)]^+ \mathrm{vec}(Q_1^* M_0^* Q_2^*),$$

where M_0^* is defined in (19), $\Gamma^* \in \mathbb{R}^{pH \times pH}$, $\widehat{Q}_1^*$, and $\widehat{Q}_2^*$ are the eigenprojectors associated with the smallest eigenvalues of $M_0^* M_0^{*T}$ and $M_0^{*T} M_0^*$. As Proposition 7, the following one is an easy application of Theorem 4.

Proposition 9. *If (1), (2), (19), $\widehat{M} \xrightarrow{\text{a.s.}} M_0$ and $\Gamma^* \xrightarrow{\mathbb{P}} \Gamma$ hold, then the test described in (4) with the statistic $\widehat{\Lambda}_2$ and calculation of quantile with Λ_2^* is consistent.*

[7] We write the expression of $\widehat{\Lambda}_2$ another way for the reasons explained in Remark 2 but one can recover the original expression by noting that for any symmetric matrix A, $A^+ H = (AH)^+$ if H is an orthonormal basis of a vector subspace of $\mathrm{Im}(A)$.

3.3 Minimum Discrepancy approach

Noting that $\{\text{rank}(M) = m\}$ has co-dimension $(H-m)(p-m)$ and applying (12) we get the following proposition[8].

Proposition 10. *If (1), (2), and (3) hold, we have*

$$\widehat{\Lambda}_3 \xrightarrow{\text{d}} \chi^2_{(H-m)(p-m)}.$$

In general a minimizer

$$\widehat{M}_c = \underset{\text{rank}(M)=m}{\text{argmin}} \ \text{vec}(\widehat{M} - M)^T \widehat{\Gamma}^{-1} \text{vec}(\widehat{M} - M)$$

does not have an explicit form as it was for the constrained matrix associated with $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$. Therefore, we define

$$M_0^* = \widehat{M}_c + n^{-1/2} W^* \qquad \text{with} \quad W^* | \widehat{P} \overset{\text{d}}{\to} W \quad \text{a.s.}, \tag{20}$$

where W is defined in (1). We also define the associated CS bootstrap statistic

$$\Lambda_3^* = n \min_{\text{rank}(M)=m} \text{vec}(M_0^* - M)^T \Gamma^{*-1} \text{vec}(M_0^* - M),$$

and applying Theorem 4 we have the following result.

Proposition 11. *If (1), (2), (3), (20), $\Gamma^* \overset{\mathbb{P}}{\to} \Gamma$, and $\widehat{M} \overset{\text{a.s.}}{\to} M_0$ hold, then the test described in (4) with the statistic $\widehat{\Lambda}_3$ and calculation of quantiles with Λ_3^* is consistent.*

Remark 3. The set of assumptions needed to obtain Proposition 10 is stronger than the ones stated in propositions 6 and 8 ensuring the convergence of $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$. As a consequence this is also true for Proposition 11 with respect to propositions 7 and 9. The main difference is that we add the assumption on Γ to be non deficient. This assumption cannot be alleviated in the statement but is not as restrictive in practice. On the first hand, if Γ is deficient the optimization under constraint has a free coordinate which implies the non-convergence of the minimizer. On the other hand, because of the semi-definite character of Γ the projection of $\widehat{M}$ on the null space of Γ is null. Then one can apply the proposition to the restriction of $\widehat{M}$ on the range of Γ. This is the case in the application to SDR in Section 4.

Remark 4. Unlike the situation of $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$, an optimization algorithm is needed to obtain $\widehat{\Lambda}_3$ and Λ_3^*, this point out an important issue of such a procedure. In [10], the authors noticed that

$$\widehat{\Lambda}_3 = n \min_{A \in H_d, B \in \mathbb{R}^{d \times l}} (\text{vec}(\widehat{M}) - \text{vec}(AB))^T \widehat{\Gamma}^{-1} (\text{vec}(\widehat{M}) - \text{vec}(AB))$$

where H_d is the set of orthogonal basis lying in $\mathbb{R}^p$ with dimension d. We follow their algorithm in the computation of $\widehat{\Lambda}_3$ (see [10], Section 3.3 for the details).

[8] See [12] for the original proof.

3.4 The statistics $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, $\widehat{\Lambda}_3$ through an example

In the introduction, we already mentioned several drawbacks and advantages of the use of $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, or $\widehat{\Lambda}_3$. The remark relied on both pivotality of the statistics and large matrix inversion. Here we develop another point of view related to the algebraic nature of the statistics. Facing the representation provided by Table 1, each statistic $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$ evaluates a different distance between $\widehat{M}$ and $\widehat{M}_c$. The first one is the distance that is optimized, but the second is another one. This has raised the issue we present here through the following example. For the sake of clarity, we consider

$$\widehat{M} = \begin{pmatrix} \widehat{\lambda}_1 & 0 \\ 0 & \widehat{\lambda}_2 \end{pmatrix} \qquad \text{with } \widehat{\lambda}_k = \frac{1}{n}\sum_{i=1}^{n} \lambda_{k,i}, \text{ for } k = 1, 2, \text{ and } (\lambda_{k,i})_{k,i} \text{ i.i.d.},$$

and we test $H_0 : d_0 = 1$ against $H_1 : d_0 > 1$. We assume that $\widehat{\lambda}_1 > \widehat{\lambda}_2$, we have $\widehat{\Lambda}_1 = n\widehat{\lambda}_2^2$. Otherwise, one can show that $\widehat{\Lambda}_2 = n\frac{\widehat{\lambda}_2^2}{\widehat{v}_2} + o_{\mathbb{P}}(1)$, with $\widehat{v}_k = \overline{(\lambda_k - \overline{\lambda_k})^2}$. For $\widehat{\Lambda}_3$ it is clear that the minimization can be done over the diagonal matrix $\operatorname{diag}(\lambda_1, \lambda_2)$ and one has

$$\widehat{\Lambda}_3 = n \operatorname*{argmin}_{\lambda_1\lambda_2=0} \left\{ \frac{\widehat{\lambda}_1 - \lambda_1}{\widehat{v}_1} + \frac{\widehat{\lambda}_2 - \lambda_2}{\widehat{v}_2} \right\} + o_{\mathbb{P}}(1) = n \min\left(\frac{\widehat{\lambda}_1^2}{\widehat{v}_1}, \frac{\widehat{\lambda}_2^2}{\widehat{v}_2} \right) + o_{\mathbb{P}}(1).$$

Accordingly, by Proposition 7, 9 and 11, the three tests can be summarized by

$$\begin{aligned} n\widehat{\lambda}_2^2 &\quad \text{compared to} \quad v_2\chi_1^2, \\ n\frac{\widehat{\lambda}_2^2}{\widehat{v}_2} &\quad \text{compared to} \quad \chi_2^2, \\ n\min\left(\frac{\widehat{\lambda}_1^2}{\widehat{v}_1}, \frac{\widehat{\lambda}_2^2}{\widehat{v}_2}\right) &\quad \text{compared to} \quad \chi_2^2, \end{aligned}$$

where $v_k = \operatorname{var}(\lambda_{k,1})$. Assume there is less variance on the estimate of the smallest eigenvalue, i.e. $v_1 > v_2$ such that $\frac{\widehat{\lambda}_1^2}{\widehat{v}_1} < \frac{\widehat{\lambda}_2^2}{\widehat{v}_2}$, this situation may arise when $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$ have similar values but different variances. Then to conduct the test, the statistic $\frac{\widehat{\lambda}_1^2}{\widehat{v}_1}$ is a better choice than $\frac{\widehat{\lambda}_2^2}{\widehat{v}_2}$. As a consequence, unlike $\widehat{\Lambda}_1$ and $\widehat{\Lambda}_2$, the statistic $\widehat{\Lambda}_3$ appears as a coherent choice because its associated minimization takes into account the variance of the estimation.

4 Application to sufficient dimension reduction

We focus on a particularly famous method in SDR called sliced inverse regression (SIR) which has been introduced in [21] to deal with the regression model

$$Y = f(PX, \varepsilon) \tag{21}$$

where $\varepsilon \perp\!\!\!\perp X \in \mathbb{R}^p$, $Y \in \mathbb{R}$, and P is a projector on the vector space E with dimension $d_0 < p$, called the central subspace. The objective is to estimate E. If X is elliptically distributed, then we have that $\Sigma^{-1}(\mathbb{E}[(X - \mathbb{E}[X])\psi(Y)] \in E$ with $\Sigma = \operatorname{var}(X)$, for any measurable function ψ. Accordingly, in order to recover the whole central subspace one needs to consider many functions ψ. For a given family of functions $(\psi_h)_{1\le h\le H}$ we define $\Psi = (\psi_1(Y), ..., \psi_H(Y))^T$. Under some additional conditions [22], the image of the matrix $\Sigma^{-1/2}\operatorname{cov}(X, \Psi(Y))$ is equal to $\Sigma^{1/2}E$. Then one can make the svd of an estimator of this matrix to obtain d_0 vectors that form an estimated basis of $\Sigma^{1/2}E$. Motivated by the curse of dimensionality, the estimation of d_0 is one of the most crucial points in SDR. To make that possible, a popular way consists in estimating the rank of $\Sigma^{-1/2}\operatorname{cov}(X,\Psi)$ using the hypothesis testing framework given by (4) (see for example [21], [8] and [10]). Since we are interested in estimating the rank, we prefer to deal directly with $\operatorname{cov}(X,\Psi)$ to avoid the introduction of an additional noise due to the estimation of the matrix Σ. Assume that $((X_1,Y_1),\cdots,(X_n,Y_n))$ is a i.i.d. sequence from model (21), denote by $\widehat{P}$ its associated empirical c.d.f. and define the quantity

$$C=\mathbb{E}[K],\qquad \text{with } K=(X-\mathbb{E}[X])(\Psi(Y)-\mathbb{E}[\Psi(Y)])^T,$$

associated with its empirical estimator

$$\widehat{C}=\overline{\widehat{K}},\qquad \text{with } \widehat{K}_i=(X_i-\overline{X})(\Psi_i-\overline{\Psi})^T,\quad \text{and } \Psi_i=\Psi(Y_i).$$

We apply the CS bootstrap to calculate the quantiles of each statistic. Facing (19) and (20), we use an independent weighted bootstrap to reproduce the asymptotic law of $\sqrt{n}(\widehat{C}-C)$, that is we define the bootstrap matrix

$$C^*=\widehat{C}_c+\overline{K^*},\quad \text{with}\quad K_i^*=w_i(\widehat{K}_i-\overline{\widehat{K}}) \tag{22}$$

where $\widehat{C}_c$ stands for the solution of an optimization problem depending on the selected statistic Λ_1, Λ_2 or Λ_3 (see Section 3 for the details) and (w_i) is a sequence of i.i.d. random variables. We also define

$$V=\operatorname{var}(\operatorname{vec}(K))\qquad \text{and}\qquad V^*=\frac{1}{n}\sum_{i=1}^{n}\operatorname{vec}(K_i^*-\overline{K^*})\operatorname{vec}(K_i^*-\overline{K^*})^T.$$

To apply propositions 7, 9, and 11, we need the following result which is of particular interest since it provides a new bootstrap procedure for SIR that is different than the one proposed in [3].

Proposition 12. *Assume that $\mathbb{E}[\|X\|^2]<+\infty$, $\mathbb{E}[\|\Psi(Y)\|^2]$ and $\mathbb{E}[\|K\|_F^4]$ are finites, if moreover (w_i) is a i.i.d. sequence of real random variables with mean 0 and variance 1, then we have*

$$\mathcal{L}_\infty(n^{1/2}\ \overline{K^*}|\widehat{P})=\mathcal{L}_\infty(n^{1/2}(\widehat{C}-C))\quad \textit{a.s. and}\quad V^*\xrightarrow{\mathbb{P}}V\quad \textit{conditionally a.s.}.$$

Remark 5. Taking a partition $\{I(h),\ h=1,\dots,H\}$ of the range of Y we recover the original SIR method with the family formed by the $p_h^{-1/2}\mathbb{1}_{\{Y\in I(h)\}}$'s with $p_h=\mathbb{P}(Y\in I(h))$. Then $C_{\text{SIR}}=\Sigma^{-1/2}\operatorname{cov}(X,\mathbb{1})D^{-1/2}$ with $\mathbb{1}=(\mathbb{1}_{\{Y_i\in I(1)\}},\dots,\mathbb{1}_{\{Y_i\in I(H)\}})^T$ and $D=\operatorname{diag}(p_h)$, is estimated by $\widehat{C}_{\text{SIR}}=\widehat{\Sigma}^{-1/2}\overline{(X-\overline{X})\mathbb{1}^T}\widehat{D}^{-1/2}$ with $\widehat{D}=\operatorname{diag}(\widehat{p}_h)$, $\widehat{p}_h=\overline{\mathbb{1}_{\{Y\in I(h)\}}}$, $\widehat{\Sigma}=\overline{(X-\overline{X})(X-\overline{X})^T}$. We have the expansion

$$\begin{aligned}n^{-1/2}(\widehat{C}_{\text{SIR}}-C_{\text{SIR}})&=n^{-1/2}\Sigma^{-1/2}\overline{((X-\mathbb{E}[X])\mathbb{1}^T}-\operatorname{cov}(X,\mathbb{1}))D^{-1/2}\\&\quad-\Sigma^{-1/2}n^{-1/2}(\widehat{\Sigma}^{1/2}-\Sigma^{1/2})C_{\text{SIR}}-C_{\text{SIR}}n^{-1/2}(\widehat{D}_p^{1/2}-D_p^{1/2})D_p^{-1/2}+o_{\mathbb{P}}(1).\end{aligned}$$

As a consequence, the matrix $\Sigma^{-1/2}$ and the weights p_h's are playing an important role on the asymptotic of the matrix SIR. They introduce some other terms in the asymptotic distribution and clearly the simple bootstrap presented before does not work for SIR as it was originally defined. Even if we believe that a more evolved weighted bootstrap works to bootstrap $\sqrt{n}(\widehat{C}_{\text{SIR}}-C_{\text{SIR}})$, we emphasize that it may be less accurate than the one we propose since it complicates the asymptotic without being necessary for testing the rank.

Recall that m is a non-negative integer, for $k \in \{1,2,3\}$ and $B \in \mathbb{N}^*$ we calculate independent copies $\Lambda^*_{k,1}, ..., \Lambda^*_{k,B}$ with the CS bootstrap algorithm corresponding to each statistic. Then we estimate the quantile with

$$q^*_k(\alpha) = \inf_{t\in\mathbb{R}}\{F^*_k(t) > \alpha\} = \Lambda^*_{k,(\lceil B\alpha\rceil)}, \qquad \text{where } F^*_k(t) = \frac{1}{B}\sum_{b=1}^{B} \mathbb{1}_{\{\Lambda^*_{k,b}\leq t\}},$$

$\lceil\cdot\rceil$ is the integer ceiling function and $\Lambda^*_{k,(\cdot)}$ stands for the rank statistic associated to the sample $\Lambda^*_{k,1}\dots\Lambda^*_{k,B}$. On the first hand, we conduct the test described by (4) using the CS bootstrap, i.e.

$$H_0 \text{ is rejected if} \qquad \widehat{\Lambda}_k > \widehat{q}^*_k(\alpha). \tag{23}$$

On the other hand, the traditional test is conducted by comparing the statistic $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ to the quantile of their asymptotic law respectively given by propositions 8 and 10. For $\widehat{\Lambda}_1$, in general the limit in law is quite complicated[9] (see Proposition 6), so that we use approximations: the Wood's approximation (see [24]) as it is computed in the R software, an adjusted version $\widehat{\Lambda}_{1,\text{adj.}} = \widehat{\Lambda}_1/a \overset{\text{d}}{\to} \chi^2_b$, with $a = \sum_{k=1}^s \omega^2/\sum_{k=1}^s \omega_k$, $b = (\sum_{k=1}^s \omega_k)^2/\sum_{k=1}^s \omega_k^2$, and a re-scaled version $\widehat{\Lambda}_{1,\text{sc}} = \widehat{\Lambda}_1/c \overset{\text{d}}{\to} \chi^2_s$, $c = \overline{\omega}$ (see [4] for these two corrections).

In all the simulations we compute the matrix $\widehat{C}$ by taking $\Psi(t) = (\mathbb{1}_{\{y\in I(1),\dots,y\in I(H)\}})$ where the $I(h)$'s form an equi-partition of the range of the data $Y_1,\dots,Y_n$. In the whole study we put $(p,H) = (6,5)$, $B = 1000$ and we consider $n = 50, 100, 200, 500$. Although the parameter H does not really affect the SIR method, we choose it globally good with respect to all the situations.

The first model we study is the following standard model:

$$\text{Model I:} \qquad Y = X_1 + .1e \qquad \text{with } e \perp\!\!\!\perp X, \quad X \overset{\text{d}}{=} \mathcal{N}(0,I), \quad e \overset{\text{d}}{=} \mathcal{N}(0,1).$$

In order to highlight guidelines (A) and (B), we produce in figure 1 two graphics each representing situation under H_1 and H_0 for the statistic $\widehat{\Lambda}_3$. Similar graphics dealing with $\widehat{\Lambda}_2$ have been drawn but are not presented here. On the first one we see that even if the sample is under H_1 the bootstrap distribution reflects H_0. As a consequence, guideline (A) is satisfied and the power of the bootstrap test is going to 1. The second graph shows that the statistic distribution is closer to the bootstrap distribution than its asymptotic distribution. This has no reason to occur when the statistic is not pivotal (see the introduction and [15] for the details). As a consequence, we believe that this good fitting is due to Guideline B.

In figure 2 we analyse the asymptotic distribution of $\widehat{q}(\alpha)$ in model I for each statistic. To measure the error we consider the behaviour of

$$F_n(\widehat{q}(\alpha)),$$

which is optimally equal to $1-\alpha$. To make that possible, F_n is estimated with a large sample size so that the estimation error is negligible. Then we run over 100 samples the CS bootstrap to provide, for each sample, a bootstrap estimation of the quantile $\widehat{q}(\alpha)$. The associated boxplot for $n = 100$, 200, 500 are provided in Figure 2. As a consequence, we may notice that the behaviour of $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ are quite similar facing the one of $\widehat{\Lambda}_1$. Even if every boxplot argues

[9] When the predictors are normally distributed, it has been shown that $\widehat{\Lambda}_1$ is asymptotically chi-squared distributed (see [8]). The authors also pointed out that it was less robust than the weighted chi-squared asymptotic as soon as the predictors distribution deviates from normality. As a result, we keep in the nonparametric framework by avoiding such asymptotic in this simulation study.

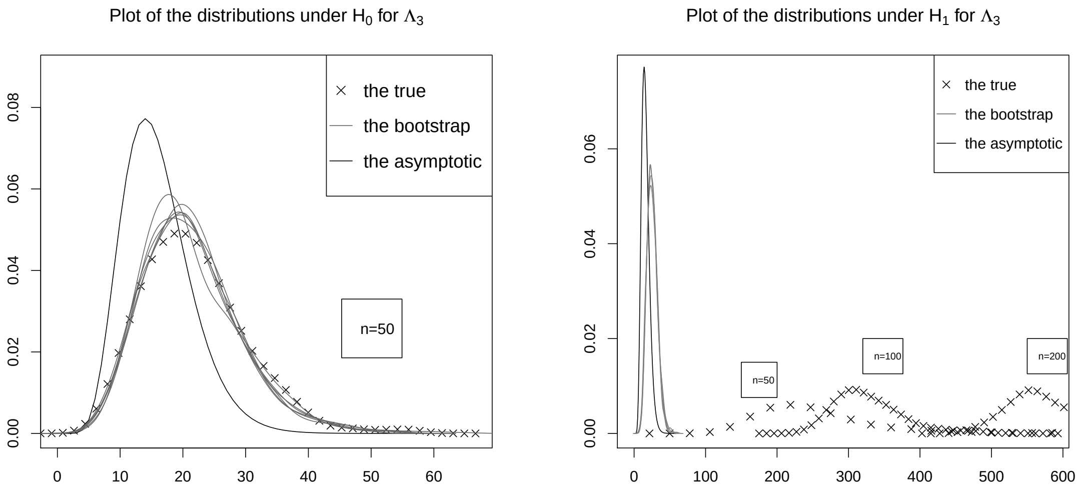

Figure 1: Plot of the asymptotic distribution, and the estimated distribution of the statistic and the bootstrap statistic for $\widehat{\Lambda}_3$ in the case of Model I.

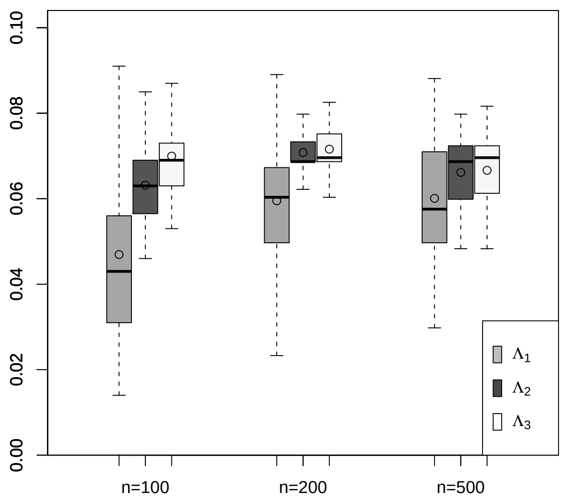

Figure 2: Bowplot over 100 samples of $\widehat{q}(\alpha)$ for $\widehat{\Lambda}_1$, $\widehat{\Lambda}_2$, $\widehat{\Lambda}_3$ and $\alpha = 0.95$ in the case of Model I for different values of n .

n	m	$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
		Wood	Resc.	Adj.	CB $\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	CB $\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$	CB $\widehat{\Lambda}_3$
50	0	0.9988	0.9998	0.9988	0.9988	1.0000	1.0000	1.0000	1.0000
	1	0.0326	0.0590	0.0336	0.0494	0.3466	0.0744	0.3098	0.07
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0386	0.052	0.0388	0.0456	0.1494	0.0676	0.1466	0.0722
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0474	0.055	0.0476	0.0514	0.096	0.0646	0.0954	0.0664
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0492	0.0514	0.0494	0.0516	0.0656	0.0584	0.0654	0.0584

Table 2: Estimated levels and power in Model I for $\alpha = 5\%$.

n	m	$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
		Wood	Resc.	Adj.	CB $\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	CB $\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$	CB $\widehat{\Lambda}_3$
50	0	0.9646	0.9928	0.9656	0.9682	1.0000	1.0000	1.0000	1.0000
	1	0.0318	0.0628	0.0324	0.0496	0.3412	0.0588	0.3042	0.0628
100	0	0.9996	1.0000	0.9996	0.9996	1.0000	1.0000	1.0000	1.0000
	1	0.0336	0.0486	0.0344	0.0412	0.1516	0.0696	0.1432	0.0718
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0378	0.0486	0.038	0.0424	0.0844	0.0602	0.0832	0.0604
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0454	0.0502	0.0458	0.0474	0.0638	0.0606	0.0634	0.0608

Table 3: Estimated levels and power in Model Ia for $\alpha = 5\%$.

for convergence to $1-\alpha$, testing with $\widehat{\Lambda}_1$ seems a better choice when n is small because of a quasi immediate convergence of the bias. When n increase, this is no longer evident because the variance of either $\widehat{\Lambda}_2^*$ or$\widehat{\Lambda}_3^*$ is smaller.

Furthermore, we go into details in Table 2 by running Model I over 5000 samples. For each of them and every statistic, we conduct the bootstrap test (23) and its traditional version. The table presents for each $m \leq d_0$, the proportion of rejected tests. This corresponds to either estimate of the power or estimate of the level.

Although it has not the best power, the clear winner is the tests based on $\widehat{\Lambda}_1$. Inside this group, for any sample number, the bootstrap and the rescaled version are the closest to the nominal level. Concerning $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$ the result are quite impressive when n is small: for $n = 100$, whereas traditional testing makes a type I error 30% of the time, the bootstrap testing goes wrong around 7%. This confirms observation on the second graph of Figure 1.

n	m	$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
		Wood	Resc.	Adj.	CB $\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	CB $\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$	CB $\widehat{\Lambda}_3$
50	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.034	0.1072	0.034	0.0378	0.2122	0.0396	0.1394	0.015
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.037	0.0904	0.0374	0.0404	0.0986	0.0572	0.0614	0.0284
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0484	0.096	0.0488	0.0518	0.0708	0.066	0.056	0.0506
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0486	0.0912	0.0486	0.0490	0.0598	0.0664	0.0612	0.0674

Table 4: Estimated levels and power in Model Ib for $\alpha = 5\%$.

n	m	$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
		Wood	Resc.	Adj.	CB $\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	CB $\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$	CB $\widehat{\Lambda}_3$
50	0	0.9308	0.9884	0.9428	0.9448	1.0000		0.9988 1.0000	0.9988
	1	0.0036	0.0148	0.0050	0.0086	0.1816	0.0148	0.1404	0.0130
100	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0072	0.0122	0.0082	0.0096	0.0536	0.02	0.0496	0.021
200	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0076	0.0114	0.0086	0.0102	0.0252	0.0192	0.0248	0.02
500	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	1	0.0068	0.0076	0.007	0.0082	0.012	0.011	0.012	0.011

Table 5: Estimated levels and power in Model II for $\alpha = 1\%$.

In Table 3 and Table 4 we consider the same model than Model I excepted that we change the distribution of the predictors: in Model Ia, X has independent coordinates with a student distribution with 5 degrees of freedom, in Model Ib, $X \stackrel{\mathrm{d}}{=} .1X_1\epsilon + X_2(1-\epsilon)$ with $\epsilon \stackrel{\mathrm{d}}{=} \mathcal{B}(1/2)$, $X_1 \stackrel{\mathrm{d}}{=} \mathcal{N}((6,0,\cdots,0),I)$, $X_2 \stackrel{\mathrm{d}}{=} \mathcal{N}(0,I)$. For this two models, we have similar conclusions than model I with two new things. First, the rescaled version is not robust to the distribution of the predictors (Table 4). Second, the algorithm employed to optimized $\widehat{\Lambda}_3$ could failed at very small sample size.

We introduce a non linear relationship by considering the model

$$\text{Model II:} \qquad Y = \tanh(X_1) + .1e \qquad \text{with } e \perp\!\!\!\perp X, \quad X \stackrel{\mathrm{d}}{=} \mathcal{N}(0,I), \quad e \stackrel{\mathrm{d}}{=} \mathcal{N}(0,1).$$

In Table 5, we present similar results as in tables 3-5 with the difference that the nominal level is $\alpha = 1\%$ in order to highlight differences in the power of each test. Again, the CS bootstrap induces a large improvement of the accuracy of the test with $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. At $n = 50$, the test based on $\widehat{\Lambda}_1$ is less powerful than the others but it is more accurate under H_0. The winner remains the CS bootstrap with $\widehat{\Lambda}_1$. A new important things is that at $n = 500$, it seems better to use the CS bootstrap with $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. Actually this is due to the variance of the formers which is smaller than the variance of Λ_1^* as it was already highlighted in Figure 2.

We conclude by increasing difficulty considering the following model, introduced in [21],

$$\text{Model III:} \qquad Y = \frac{X_1}{.5 + (X_2+2)^2} + e \qquad e \perp\!\!\!\perp X, \qquad X \stackrel{\mathrm{d}}{=} \mathcal{N}(0,I)$$

We still present in Table 6 the estimated level and power with the nominal level $\alpha = 2\%$ for each test. For such a model the conclusions are quite mitigated because it induces a trade-off between high power and accurate level. Indeed when n is small, the better powers are provided by the traditional tests with $\widehat{\Lambda}_2$ and $\widehat{\Lambda}_3$. Nevertheless the more accurate levels can be found looking at the CS bootstrap with $\widehat{\Lambda}_2$ (n = 100) or $\widehat{\Lambda}_1$ (n = 200). Moreover the tests associated to $\widehat{\Lambda}_1$ without bootstrap are the worst concerning this model. Accordingly, the simulation study highlighted the good behaviour of the CS bootstrap: in every model it improves the accuracy of the traditional test for each statistic. One may remember that the bias of the CS bootstrap with $\widehat{\Lambda}_1$ has the faster rate of convergence with respect to the CS bootstrap of $\widehat{\Lambda}_2$ or $\widehat{\Lambda}_3$. Otherwise, the variance of $\widehat{\Lambda}_1^*$ may be greater than the variance of $\widehat{\Lambda}_2^*$ or $\widehat{\Lambda}_3^*$. Finally, for the simple models it seems better to use the CS bootstrap with the statistic $\widehat{\Lambda}_1$.

5 Concluding remarks

Along this study, we found that the main advantages of the CS bootstrap are:

n	m	$\widehat{\Lambda}_1$				$\widehat{\Lambda}_2$		$\widehat{\Lambda}_3$	
		Wood	Resc.	Adj.	CB $\widehat{\Lambda}_1$	$\widehat{\Lambda}_2$	CB $\widehat{\Lambda}_2$	$\widehat{\Lambda}_3$	CB $\widehat{\Lambda}_3$
	0	0.9950	0.9992	0.9962	0.9960	1.0000	0.9966	1.0000	0.9966
50	1	0.3750	0.5342	0.3990	0.4676	0.9074	0.5066	0.8344	0.3270
	2	0.0078	0.0156	0.0086	0.0240	0.0620	0.0164	0.0344	0.0136
	0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
100	1	0.9330	0.9556	0.9368	0.9446	0.9952	0.9842	0.9934	0.9806
	2	0.0134	0.0176	0.0138	0.0210	0.0306	0.0228	0.0266	0.0278
	0	1.000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000
200	1	1.000	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000
	2	0.0154	0.0182	0.0158	0.0198	0.025	0.024	0.0244	0.026
	0	1.0000	1.000	1.0000	1.0000	1.0000	1.000	1.0000	1.0000
500	1	1.0000	1.000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	2	0.0184	0.0194	0.0184	0.02	0.0228	0.0228	0.0228	0.023

Table 6: Estimated levels and power in Model II for $\alpha = 2\%$.

1. Alternative to the asymptotic comparison. This argument is even stronger since the asymptotic law can be unknown (or difficult to estimate) or the asymptotic law remains too much different from the statistic law (e.g. large matrix inversion).

2. By Theorem 4, which provides its consistency, the CS bootstrap works under mild assumptions. Essentially, we ask the manifold to be locally smooth, and we require a bootstrap 18 of the unconstrained estimator.

3. The CS bootstrap is computationally as simple than the considered statistic.

4. In the case of rank testing, the CS bootstrap clearly improves the accuracy of traditional testing (cf. the simulation study).

Besides, there exists some natural extensions of the previous work. First although it is suitable for testing, the form of the objective function bQ is quiet restrictive. For example, we believe that the CS bootstrap could be extended to M and Z estimation. Secondly, conditions that guarantee

$$\widehat{q}(\alpha) = q_n(\alpha) + o_{\mathbb{P}}(n^{-1/2})$$

have not been provided yet. This would valid theoretically the use of the CS bootstrap with respect to traditional testing.

Appendix

Proof of Lemma 1

The whole proof is made conditionally on the sample. By definition of $\widehat{\theta}_c$, with high probability, A^* is full rank for n large enough, we have

$$\|A^{*1/2}(\theta_c^* - \theta_c)\| \leq \|A^{*1/2}(\theta_c^* - \theta_0^*)\| + \|A^{*1/2}(\theta_0^* - \theta_c)\| \leq 2\|A^{*1/2}(\theta_0^* - \theta_c)\|. \quad (24)$$

Then since $\theta_0^* - \widehat{\theta}_c \xrightarrow{\mathbb{P}} 0$, $\widehat{\theta}_c \to \theta_c$ and because $A^* \xrightarrow{\mathbb{P}} A$ is full rank, one gets that $\theta_c^* \xrightarrow{\mathbb{P}} \theta_c$. Therefore, since θ_c is $\mathcal{M}$-nonsingular and reffering to Definition 1, we get

$$\operatorname*{argmin}_{\theta\in\mathcal{M}} \|\Gamma^{*1/2}(\theta_0^* - \theta)\| = \operatorname*{argmin}_{g(\theta)=0} \|\Gamma^{*1/2}(\theta_0^* - \theta)\|,$$

with g continuously differentiable on θ_c and $J_g(\theta_c)$ full rank. By assumption on g, θ_c^*, at least for n large enough, satisfies the first order conditions, that are

$$\begin{cases} A^*(\theta_0^* - \theta_c^*) - J_g^T(\theta_c^*)\lambda_n^* = 0 \\ g(\theta_c^*) = 0 \end{cases}$$

where λ_n^* is the Lagrange multiplier. Using a Taylor expansion of g around $\widehat{\theta}_c$, we get $g(\theta_c^*) = g(\widehat{\theta}_c) + J_g^T(\widehat{\theta}_c)(\theta_c^* - \widehat{\theta}_c) + o_{\mathbb{P}}(\|\theta_c^* - \widehat{\theta}_c\|)$, and with the previous equations we have

$$\begin{pmatrix} A^* & J_g^T(\theta_c^*) \\ J_g(\widehat{\theta}_c) & 0 \end{pmatrix} \begin{pmatrix} \theta_c^* - \widehat{\theta}_c \\ \lambda_n^* \end{pmatrix} = \begin{pmatrix} A^*(\theta_0^* - \widehat{\theta}_c) \\ o_{\mathbb{P}}(\|\theta_c^* - \widehat{\theta}_c\|) \end{pmatrix}.$$

Now by Slutsky's lemma, we get

$$\begin{pmatrix} A & J_g^T(\theta_c) \\ J_g(\theta_c) & 0 \end{pmatrix} \begin{pmatrix} n^{1/2}(\theta_c^* - \widehat{\theta}_c) \\ n^{1/2}\lambda_n^* \end{pmatrix} = n^{1/2} \begin{pmatrix} A(\theta_0^* - \widehat{\theta}_c) \\ 0 \end{pmatrix} + o_{\mathbb{P}}(1),$$

and the conclusion follows by multiplying on the left by the matrix

$$\left(A^{-1} - PA^{-1}, \quad A^{-1}J_g^T(\theta_c)(J_g(\theta_c)A^{-1}J_g^T(\theta_c))^{-1}\right)$$

with $P = A^{-1}J_g^T(\theta_c)(J_g(\theta_c)A^{-1}J_g^T(\theta_c))^{-1}J_g(\theta_c)$.

□

Proof of Theorem 4

The proof is divided in two parts each corresponding to the level and the power of the test. Assume H_0 and define F_n and F_∞ respectively as the c.d.f. of $\widehat{\Lambda}$ and the weak limit of F_n. Note that we can apply Proposition 2 to get

$$n^{1/2} \begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{\theta}_c - \theta_0 \end{pmatrix} = n^{1/2} \begin{pmatrix} I \\ I - P \end{pmatrix} (\widehat{\theta} - \theta_0) + o_{\mathbb{P}}(1),$$

and Theorem 3 to get conditionally a.s.

$$n^{1/2} \begin{pmatrix} \theta_0^* - \widehat{\theta}_c \\ \theta_c^* - \widehat{\theta}_c \end{pmatrix} = n^{1/2} \begin{pmatrix} I \\ I - P \end{pmatrix} (\theta_0^* - \widehat{\theta}_c) + o_{\mathbb{P}}(1).$$

with P detailed in the statement of Proposition 2. Using (11), (18) and Slutsky's theorem we have

$$\mathcal{L}_\infty(\Lambda^*|\widehat{P}) = \mathcal{L}_\infty(\widehat{\Lambda}) \quad \text{a. s.}.$$

In other words, with probability 1, $\widehat{F}$ converges pointwise to F_∞. As in [23] chapter 23, Lemma 3, consider Δ the set of discontinuity of F_∞^{-1}. For every $\alpha \in]0,1[\backslash\Delta$, we have $\widehat{q}(\alpha) \longrightarrow q(\alpha)$ a.s. (see for instance [23], chapter 21). Using Slutsky's theorem, we get $\mathcal{L}_\infty(\widehat{\Lambda} - \widehat{q}(\alpha)) = \mathcal{L}_\infty(\widehat{\Lambda} - q(\alpha))$, accordingly

$$\mathbb{P}(\widehat{\Lambda} \le \widehat{q}(\alpha)) \longrightarrow F_\infty(q(\alpha)) \quad \text{for all } \alpha \in]0,1[\backslash\Delta.$$

Because F_∞ is continuous $F_\infty(q(\alpha)) = \alpha$. Since F_∞ is non-decreasing, Δ is denumerable, since $\alpha \mapsto \mathbb{P}(\widehat{\Lambda} \le \widehat{q}(\alpha))$ is non-decreasing with continuous limit, the convergence is uniform and so holds for every $\alpha \in]0,1[$. This concludes the proof for the level. It remains to show that the power of the test goes to 1. Assume H_1 and let $\alpha \in]0,1[$, the statistic $\widehat{\Lambda}$ goes to infinity in probability and it suffices to show that with probability 1 the bootstrap quantile $\widehat{q}(\alpha)$ remains bounded. This means exactly that conditionally a.s. the sequence Λ^* is tight. Note that conditionally a.s. we have

$$\Lambda^* \leq n\|A^{*1/2}(\widehat{\theta}_c - \theta_0^*)\|^2 = \widetilde{\Lambda}^*,$$

where $\widetilde{\Lambda}^*$ converges in distribution by (17), and is therefore tight. □

Proof of Proposition 6

We have

$$\widehat{\Lambda}_1 = \|n^{1/2}\widehat{Q}_1\widehat{M}\widehat{Q}_2\|_F^2 = \|n^{1/2}\operatorname{vec}(\widehat{Q}_1\widehat{M}\widehat{Q}_2)\|^2. \tag{25}$$

By the Delta method and because H_0 is realized, we can apply convergence results about eigenprojectors to both matrices $\widehat{M}^T\widehat{M}$ and $\widehat{M}\widehat{M}^T$ to obtain the $\sqrt{n}$-convergence for $\widehat{Q}_1$ and $\widehat{Q}_2$. Then we write

$$\begin{aligned} n^{1/2}\widehat{Q}_1\widehat{M}\widehat{Q}_2 &= n^{1/2}\widehat{Q}_1(\widehat{M}-M)\widehat{Q}_2 + n^{1/2}(\widehat{Q}_1 - Q_1)M(\widehat{Q}_2 - Q_2) \\ &= n^{1/2}Q_1(\widehat{M}-M)Q_2 + O_{\mathbb{P}}(n^{-1/2}), \end{aligned}$$

which suffices to obtained the first statement of the theorem. For the second statement, the symmetric matrix $(Q_2 \otimes Q_1)\Gamma(Q_2 \otimes Q_1)$ is estimated consistently by $(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)$ and so are its eigenvalues. □

Proof of Proposition 8

We can notice that $\sqrt{n}\widehat{Q}_1\widehat{M}\widehat{Q}_2$ has the same asymptotic law than $\sqrt{n}Q_1(\widehat{M}-M)Q_2$ whose asymptotic variance is consistently estimated by $[(\widehat{Q}_2 \otimes \widehat{Q}_1)\widehat{\Gamma}(\widehat{Q}_2 \otimes \widehat{Q}_1)]^+$ (see the proof of Proposition 6). □

Proof of Proposition 12

Recall that $\widehat{K}_i = (X_i - \overline{X})(\Psi_i - \overline{\Psi})$, $K_i^* = w_i(\widehat{K}_i - \overline{\widehat{K}})$ and define $K_i = (X_i - \mathbb{E}[X])(\Psi_i - \mathbb{E}[\Psi])$. First note that, by Slutsky's theorem, $\sqrt{n}\ \overline{K^*}$ has the same asymptotic law than $n^{-1/2}\sum_{i=1}^n w_i(\widehat{K}_i - \mathbb{E}[K])$. Then we can develop

$$\begin{aligned} & n^{-1/2}\sum_{i=1}^n w_i(\widehat{K}_i - \mathbb{E}[K]) \\ &= n^{-1/2}\sum_{i=1}^n w_i((X_i - \mathbb{E}[X])(\Psi_i - \overline{\Psi})^T - \mathbb{E}[K]) + (\mathbb{E}[X] - \overline{X})n^{-1/2}\sum_{i=1}^n w_i(\Psi_i - \overline{\Psi})^T \\ &= n^{-1/2}\sum_{i=1}^n w_i(K_i - \mathbb{E}[K]) + n^{-1/2}\sum_{i=1}^n w_i(X_i - \mathbb{E}[X])(\mathbb{E}[\Psi] - \overline{\Psi})^T \\ &\qquad + (\mathbb{E}[X] - \overline{X})n^{-1/2}\sum_{i=1}^n w_i(\Psi_i - \overline{\Psi})^T. \end{aligned}$$

Checking a Lindeberg condition as bellow to ensure the weak convergence of $n^{-1/2}\sum_{i=1}^{n} w_i(X_i - \mathbb{E}[X])$ and $n^{-1/2}\sum_{i=1}^{n} w_i(\Psi_i - \overline{\Psi})^T$, and using the Slutsky's theorem we get conditionally a.s.

$$n^{1/2}\ \overline{K^*} = n^{-1/2}\sum_{i=1}^{n} w_i(K_i - \mathbb{E}[K]) + O_{\mathbb{P}}(n^{-1/2}).$$

We can apply the multidimensional version of the Lindeberg's central limit theorem (see for instance [5], Corollary 18.2), provided that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\widehat{V}^{-1/2}w_i\xi_i\|^2\mathbb{1}_{\{\|\widehat{V}^{-1/2}w_i\xi_i\|>\nu n^{-1/2}\}}|\widehat{P}] \xrightarrow{\text{a.s.}} 0,$$

where $\xi_i = \text{vec}(K_i - \mathbb{E}[K])$ and $\widehat{V} = \frac{1}{n}\sum_{i=1}^{n}(\xi_i - \overline{\xi})(\xi_i - \overline{\xi})^T$. The above convergence is a consequence of the Lebesgue domination theorem which ensure that each term of the sum goes to 0, afterwards we can conclude by the Cesaro's Lemma. Thus we have proved that conditionally a.s.

$$n^{-1/2}\widehat{V}^{-1/2}\sum_{i=1}^{n} w_i\xi_i \xrightarrow{\text{d}} \mathcal{N}(0, I),$$

and it remains to note that $\widehat{V} \overset{\text{a.s.}}{\to} V$ the variance of the limit in law of $\sqrt{n}(\widehat{C} - C)$ provided that K has a finite order 2 moment. For the second convergence, we note that conditionally a.s.

$$V^* - \widehat{V} = \frac{1}{n}\sum_{i=1}^{n}(w_i^2 - 1)\xi_i\xi_i^T + o_{\mathbb{P}}(1),$$

then by noting v_i a coordinate of $\xi_i\xi_i^T$ we calculate

$$\mathbb{E}\left[\left(n^{-1}\sum_{i=1}^{n}(w_i^2 - 1)v_i\right)^2\right] = n^{-2}\mathbb{E}[(w_i^2 - 1)^2]\sum_{i=1}^{n} v_i^2$$

which goes to 0 a.s. provided that K has a finite order 4 moment. We conclude by using the Markov inequality to get that $V^* \overset{\mathbb{P}}{\to} \widehat{V}$ conditionally a.s.. $\square$

References

[1] Miguel A. Arcones and Evarist Giné. On the bootstrap of M-estimators and other statistical functionals. In *Exploring the limits of bootstrap (East Lansing, MI, 1990)*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., pages 13–47. Wiley, New York, 1992.

[2] Philippe Barbe and Patrice Bertail. *The weighted bootstrap*, volume 98 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1995.

[3] M. Pilar Barrios and Santiago Velilla. A bootstrap method for assessing the dimension of a general regression problem. *Statist. Probab. Lett.*, 77(3):247–255, 2007.

[4] M. Peter Bentler and Jun Xie. Corrections to test statistics in principal hessian directions. *Statist. Probab. Lett.*, 47(4):381–389, 2000.

[5] R. N. Bhattacharya and R. Ranga Rao. *Normal approximation and asymptotic expansions*. John Wiley & Sons, New York-London-Sydney, 1976. Wiley Series in Probability and Mathematical Statistics.

[6] Dennis D. Boos. On generalized score tests. *Amer. Statist.*, 46:327–333, 1990.

[7] E. Bura and J. Yang. Dimension estimation in sufficient dimension reduction: a unifying approach. *J. Multivariate Anal.*, 102(1):130–142, 2011.

[8] Efstathia Bura and R. Dennis Cook. Extending sliced inverse regression: the weighted chi-squared test. *J. Amer. Statist. Assoc.*, 96(455):996–1003, 2001.

[9] Snigdhansu Chatterjee and Arup Bose. Generalized bootstrap for estimating equations. *Ann. Statist.*, 33(1):414–436, 2005.

[10] R. Dennis Cook and Liqiang Ni. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. *J. Amer. Statist. Assoc.*, 100(470):410–428, 2005.

[11] John G. Cragg and Stephen G. Donald. On the asymptotic properties of LDU-based tests of the rank of a matrix. *J. Amer. Statist. Assoc.*, 91(435):1301–1309, 1996.

[12] John G. Cragg and Stephen G. Donald. Inferring the rank of a matrix. *J. Econometrics*, 76(1-2):223–250, 1997.

[13] Morris L. Eaton and David Tyler. The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis. *J. Multivariate Anal.*, 50(2):238–264, 1994.

[14] B. Efron. Bootstrap methods: another look at the jackknife. *Ann. Statist.*, 7(1):1–26, 1979.

[15] Peter Hall. *The bootstrap and Edgeworth expansion*. Springer Series in Statistics. Springer-Verlag, New York, 1992.

[16] Peter Hall and Brett Presnell. Intentionally biased bootstrap methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 61(1):143–158, 1999.

[17] Peter Hall and Susan R. Wilson. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 47(2):757–762, 1991.

[18] Feifang Hu and John D. Kalbfleisch. The estimating function bootstrap. *Canad. J. Statist.*, 28(3):449–499, 2000. With discussion and rejoinder by the authors.

[19] Frank Kleibergen and Richard Paap. Generalized reduced rank tests using the singular value decomposition. *J. Econometrics*, 133(1):97–126, 2006.

[20] S. Lele. Resampling using estimating equations. In *Estimating functions*, volume 7 of *Oxford Statist. Sci. Ser.*, pages 295–304. Oxford Univ. Press, New York, 1991.

[21] Ker-Chau Li. Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.*, 86(414):316–342, 1991.

[22] François Portier and Bernard Delyon. Optimal transformation: A new approach for covering the central subspace. *To appear in Journal of Multivariate Analysis*, 2012.

[23] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[24] Andrew T. A. Wood. An f approximation to the distribution of a linear combination of chi-squared variables. *Comm. Statist. Simulation and Computation*, 18(4):1439–1456, 1989.